DOCUMENT RESUME

ED 038 183 PS 003 161

AUTHOR Fleming, Elyse S.; Anttonen, Ralph C.
TITLE Teacher Expectancy or My Fair Lady.
INSTITUTION Case Western Reserve Univ., Cleveland, Ohio.
PUB DATE 5 Mar 70
NOTE 18p.; Paper presented at the annual meeting of the American Educational Research Association, Minneapolis, Minnesota, March 5, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.00
DESCRIPTORS Analysis of Variance, Cognitive Development, Elementary School Teachers, Grade 2, Intelligence Quotient, *Intelligence Tests, Socioeconomic Status, Student Improvement, *Teacher Attitudes, *Teacher Characteristics
IDENTIFIERS Self fulfilling Prophecy

ABSTRACT

This study examined the effects of teacher expectancy on changes in intelligence, known as the phenomenon of the self-fulfilling prophecy. Differences in sex, socioeconomic status, and teacher opinion about standardized tests were control factors. The teachers completed a questionnaire that provided the basis for classifying teachers into high, middle, and low opinion groups with regard to the value they placed upon the usefulness of intelligence tests. A total of 859 second grade students in 39 classes, with high and low poverty backgrounds, were randomly assigned to one of four treatment conditions. Teachers were provided with one of four kinds of test information for each group: (1) Kuhlmann-Anderson IQ reported as tested, (2) IQs inflated by 16 points, (3) Primary Mental Abilities percentiles, and (4) intelligence test information withheld entirely. Following re-testing several months later, teachers were asked to assess IQ accuracy for each child. No significant IQ differences were found among the four treatment groups or between sexes. There were significant differences between teacher opinion, socioeconomic status, and for the interaction of teacher opinion and socioeconomic status. The study failed to support a generalized self-fulfilling prophecy. (Author/DR)

ED038183

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TEACHER EXPECTANCY OR MY FAIR LADY

Elyse S. Fleming and Ralph G. Anttonen

Case Western Reserve University

Presented at the 1970 Meeting

American Educational Research Association

Minneapolis, Minnesota

March 5, 1970

PS 003161

Widespread publicity given Rosenthal's Pygmalion in the Classroom (1968) has had a far-reaching effect on educators, many of whom have accepted the findings unquestioningly. Recently the generalizability of the self-fulfilling phenomenon has been subjected to critical review. (Barber and Silver, 1968a,b; Thorndike, 1968, 1969; Gephart, 1969; Snow, 1969). The critics suggest that the data gathering methods, the techniques of analysis and the data presentation are not sufficiently rigorous to warrant the kind of generalizations Rosenthal has made. Further attempts at replication of Rosenthal's earlier experimental procedures as well as the later educational application have not demonstrated the phenomenon's universality (Barber, et. al. 1969; Claiborn, 1969).

Barber et. al. (1969) sought in five investigations to cross-validate an earlier study by Rosenthal which purported to demonstrate the experimenter bias effect. All five attempts at replication failed to support the original experiment of Rosenthal and Fode (1963). In addition, Barber and Silver (1968a) have meticulously re-analyzed 31 studies alleging to demonstrate the experimenter bias effect raising serious questions about the conditions under which the effect may be obtained, if at all, and the methodology employed for its demonstration. In a non-laboratory experiment involving 12 first grade classrooms, Claiborn (1969) employed a design similar to Rosenthal's and Jacobson's (1968). The Claiborn study differed in that it extended over a shorter

time interval and was conducted at the conclusion of the school year. He too was unable to obtain confirmatory evidence of the expectancy effect.

Supported by the Cleveland Associated Foundations the present effort was part of a larger study including standard measures of achievement, self concept, and subject matter grades. It was an attempt to examine in an ongoing school situation whether or not the self-fulfilling prophecy would operate when conventional educational measures and procedures are used. There have been strong implications that teachers are unduly influenced by the knowledge of the ability of the children in their charge and that, in turn, the pupils' performance reflects the teacher expectancy. There is a paucity of knowledge about the relationship between a teacher's attitudes toward testing, her utilization of test results, and how this may be reflected in the performance of children in her classroom. The need for examining both sex and socio-economic differences as added dimensions within the same study represents a refinement over previous work.

With the cooperation of the Research Bureau of the Cleveland Public Schools, the personnel of 22 schools volunteered to participate in a study purporting to evaluate the usefulness of various kinds of intelligence test information to teachers. Utilizing the standardized testing procedure of the Cleveland Public School System, 1,087 second grade students in 39 classrooms representing two socio-economic levels were tested with the Kuhlmann-Anderson Intelligence Test in the Fall of 1968. Whereas the original design of the study called for twenty low and twenty middle socio-

economic level classrooms with five back-up classes for each, due to normal teacher attrition and administrative misunderstanding, the final sample consisted of 39 classrooms remaining in the study from September through June of 1968-69. While a limited number of central office administrative personnel were aware of the actual purpose of the research, no principals or teachers were apprised of the real intent, nor is there any evidence that secrecy was not maintained throughout. Informal interviews with the teachers at the conclusion of the experiment as well as written statements concerning their understanding of the intent of the research revealed only one teacher who was knowledgeable about Rosenthal and suspicious of our motives. Of the 22 schools, 11 schools with 21 classrooms had high poverty indices and 11 schools with 18 classrooms had low poverty indices. Of the original 1,087 children, 859 were present for the entire experiment as 228 either moved or were absent upon the occasion of the post-test.

The teachers completed a questionnaire sampling their attitudes about standardized tests. Information about their background characteristics was also included. The questionnaire adapted from Goslin (1967) provided the basis for classifying teachers into high, middle, and low opinion groups with regard to the value they placed upon the usefulness of intelligence tests. (See Table 1 for the items and weights on Teacher Opinion Scale).

The subjects within each classroom were randomly assigned to one of four treatment conditions early in the Fall of 1968. Under the experimental design the teachers were to receive one of

four different kinds of test information for each group: traditional IQ scores, withholding of IQ information, Primary Mental Abilities percentiles, and IQ's inflated by 16 points. Following the customary system-wide radio administration of the Kuhlmann-Anderson Intelligence test to all second graders in the Cleveland Public Schools, the PMA was administered to the randomly chosen group within each class by the experimenters' graduate assistants, tests were scored and class lists prepared for return to teachers. This followed the routine procedure for reporting test results to teachers after Fall testing except that one-fourth of the class had Primary Mental Abilities percentiles reported, one-fourth of the class had no IQ information at all and one-half of the class had IQ scores reported although unbeknownst to the teachers, one-half of these children had IQ's inflated by 16 points.

The teachers were not contacted again until scheduling for standardized achievement and self-concept testing was required in February as part of the larger study, the results of which will be reported at a later time. No further intervention was involved until May 1969 when the Kuhlmann-Anderson Intelligence Test was readministered by the experimental team. The decision to avoid communication with teachers was an attempt to insure that the study would simulate as far as possible usual teacher behavior with regard to the uses made of IQ information. In addition, a further control for a possible Hawthorne effect was introduced by re-testing in May two classes of second grade children not previously involved in the study but for whom the Cleveland Public Schools had scores from

the Fall radio-administered Kuhlmann-Anderson. These two classes were comparable to the two socio-economic levels represented in the study. All intelligence re-testing was completed by advanced graduate students who were unaware either of the purposes of the research or the group designation of the individual students. At no point were the principal investigators involved in test administration, classroom visitation or conversation with the teachers and pupils involved.

Teachers were again asked to complete a shortened form of the original attitude instrument, to state the purposes of the research as they understood them, and to rate the accuracy and usefulness of the test information provided them in the Fall for each child.

In order to test the hypothesis that greater gains in IQ would be made by children who had inflated IQ scores reported to their teachers as compared with children not so treated, a four-factor crossed analysis of covariance was performed. The 4 X 2 X 3 X 2 design included the four treatment conditions: inflated IQ, PMA percentiles, no information, traditional IQ; sex of student; the teacher opinion dimension: high, middle, low; and socio-economic level: middle and low. The dependent measure involved in the analysis was the IQ change score Fall to Spring with the covariate the pre-measure, the Fall IQ score.

The results of the overall analysis (See Table 2) reveal no significant differences among the four treatment groups. It may be noted that the overall unadjusted means for the four treatments ranged from 6.13 to 7.34 with a Grand Mean of 6.88. (See Table 3) In order to examine a possible Hawthorne effect, as stated earlier, two additional classes not previously involved in the study were

ED 036161

tested in the spring to determine expected IQ change for second graders in the Cleveland Public Schools. The overall mean for these two schools was 6.59 suggesting comparability to the gain scores attained by pupils in the experimental groups.

A further examination of the data in Table 2 shows that significant differences at the .01 level were obtained for both the teacher opinion (F = 7.04) and socio-economic factors (F=14.22). Significant differences at the .05 level were also found for the interaction of teacher opinion by socio-economic status (F = 4.30). It should be noted that because the teacher opinion dimension was defined across socio-economic status, not within, the distribution is not proportional as 11 out of the 18 teachers working in middle socio-economic status areas were in the middle opinion dimension whereas only two were in the low opinion group. In contrast, those teaching in high poverty schools tended to distribute evenly across the three opinion dimensions. Therefore, this interaction may have been contaminated by disproportionality. Recognizing this limitation one observes that overall, children in middle socio-economic schools, irrespective of treatment conditions, experienced greater gains than children in high poverty index schools, with the difference more marked within the high teacher opinion dimension. (See Tables 3 and 4 and Figure 1) It may also be observed that classes with high opinion teachers had greater overall gains in IQ than did the classes of middle opinion teachers followed, in turn, by the classes of low opinion teachers. (See Tables 3 and 4 and Figure 1) This trend was more apparent for the middle socio-economic level.

While the interpretation of these findings is not completely clear at this time, an analysis of the characteristics of the teachers in each group which is presently underway may provide an explanation.

In addition to the analysis of IQ change scores, teachers were asked to assess the accuracy of the various types of test information for their students. When an overall analysis of variance within the basic design was performed, (See Table 5) significant differences were obtained among the three IQ treatment conditions, eliminating the no information group, ($F = 4.03$, $p < .05$). Upon further examination of the significant treatment differences by means of a Scheffe[1] test, it was found that the inflated IQ group was seen as less accurate than the traditionally reported group.

It appears that in the real world of the teacher using IQ test information, the self-fulfilling prophecy does not operate as Rosenthal hypothesizes. We can only conclude that teachers are more sensitive to the functioning level of students than previously believed. Day to day living with the academic performance and behavior of children, at least for this group of teachers, provides more input than the results of an intelligence test administered on one given day. By the same token, it should be noted that the group for whom no information was provided showed no significant differences in IQ change when compared with the other three treatments. This raises some questions about the usefulness of administering group IQ tests in the primary grades, for they may not really assist in the instructional program except insofar as the teacher feels more comfortable for having had them.

Failure to confirm a generalized self-fulfilling prophecy finds support in Barber and Silver (1968a,b) who question the generalizability of the phenomenon and who would direct future attention to isolating the conditions under which the expectancy effect may operate. The external imposition of an expectancy which creates dissonance for a teacher when she is confronted with discrepant behavior in the real world of children and tests appears not to reflect itself in improved pupil performance. However, the present data provides a possible interpretation meriting further research. It may be recalled that while the treatment condition did not bring about significant effects, the teacher opinion dimension by itself and interacting with socio-economic status did. The disproportionality of the distribution of teacher's opinion by socio-economic classification tempers the interpretation of these findings, and points to future research which would seek to isolate the factors which help to bring this about.

It is clear that the way in which teachers influence pupil behavior is a far more subtle and complex phenomenon than some have suggested and that the epilogue to the play is actually yet to be written. The body of knowledge and attitudes of teachers about testing, their personal characteristics and ways of dealing with children appear to be far more critical for pupil growth than external intervention, per se. The present study suggests that teachers assess children, reject discrepant information, and operate on the basis of previously developed attitudes toward and knowledge about children and tests. This is supported in a recent study by Pallardy (1969) who found that if first grade teachers

believed that boys would achieve as well as girls in reading, the boys did in fact perform better than boys with teachers who believed them to be less successful in reading than girls. Addressing ourselves to these questions, research presently underway and projected for the future, will attempt to reduce some of the limitations of the present study and follow up some suggested clues. More specifically, in-depth study of teacher characteristics and attitudes toward testing, the effects of classroom interaction, and possible grade level differences as they may relate to pupil growth are intended. Furthermore, research designs involving more powerful statistical analyses which permit the use of classroom means as units of analysis seem essential. Control of classroom size would provide more rigor. Perhaps the question is," Professor Higgins, How did you really do it?"

Barber, T. X., and Silver, M. J. "Fact, Fiction, and the Experimenter Bias Effect." Psychological Bulletin, 1968, 70 (6, Pt. 2), 1-29.

Barber, T. X., and Silver, M. J. "Pitfalls in Data Analysis and Interpretation: A Reply to Rosenthal. "Psychological Bulletin, 1968, 70 (6, Pt. 2), 48-62.

Barber, T. X., Forgione, A., Chaves, J. F., Calverley, D. S., McPeake, J. D., and Bowen, B. "Five Attempts to Replicate the Experimenter Bias Effect." Journal of Consulting and Clinical Psychology, 1969, 33, 1-6.

Claiborn, W. L. "Expectancy Effects in the Classroom: A Failure to Replicate." Journal of Educational Psychology, 1969, 60, 377-383.

Gephart, W. J. and Antonoplos, D. P. "The Effects of Expectancy and Other Research-Biasing Factors," Phi Delta Kappan, 1969, 50, 579-583.

Goslin, D. A. Teachers and Testing. New York: Russell Sage Foundation, 1967.

Pallardy, J. N. "What Teachers Believe, What Children Achieve." The Elementary School Journal, 1969, 69, 370-4.

Rosenthal, R. "Experimenter Expectancy and the Reassuring Nature of the Null Hypothesis Decision Procedure." Psychological Bulletin, 1968, 70 (6, Pt. 2), 30-47.

Rosenthal, R. and Fode, K. "Three Experiments in Experimenter Bias." Psychological Reports, 1963, 12, 491-511.

Rosenthal, R., and Jacobson, L. Pygmalion in the Classroom. New York: Holt, Rinehart and Winston, Inc., 1968.

Snow, R. E. "Unfinished Pygmalion," Contemporary Psychology, 1969, 14, 197-199.

Thorndike, R. L. Review of R. Rosenthal and L. Jacobson, Pygmalion in the Classroom. Educational Research Journal, 1968, 5, 708-711.

Thorndike, R. L. "But You Have to Know How to Tell Time." American Educational Research Journal, 1969, 6, #4, 692.

TABLE [1]

*TEACHER OPINION SCALE

ITEM NO.	ITEM	WEIGHTS
20.	How accurate do you feel most standardized intelligence or aptitude tests are in measuring a student's potential?	
	1. Much better than other measures of ability (e.g., teacher evaluations, nonstandardized tests)	5
	2. Slightly better than other measures	4
	3. No better than other measures	3
	4. Not so good as other measures	2
	5. Much worse than other measures	1
	6. No opinion	3
21.	Which one of the following kinds of information do you feel provides the single most accurate measure of a student's intellectual ability?	
	1. Grade average	0
	2. Parents' opinions	0
	3. Standardized achievement-test scores	0
	4. IQ or scholastic aptitude-test scores	5
	5. Teacher recommendations	0
	6. The student's own opinion	0
	7. Peers' opinions	0
84.	How important do you feel the kind of intelligence measured by standardized tests is for success in school?	
	1. No opinion	3
	2. It is not important at all.	1
	3. It is less important than most other qualities.	2
	4. It is about the same in importance as most other qualities.	3
	5. It is more important than most other qualities.	4
	6. It is the most important quality for success.	5
87.	Do you think standardized intelligence tests measure primarily the intelligence people are born with, or what they have learned?	
	1. IQ tests measure only inborn intelligence.	5
	2. IQ tests measure mostly inborn intelligence but learning makes some difference.	4
	3. IQ tests measure inborn intelligence and learning about equally.	3
	4. IQ tests measure mostly learned knowledge, but inborn intelligence makes some difference.	2
	5. IQ tests measure only learned knowledge.	1
	6. No opinion.	3

*Adapted from Goslin, David A. Teachers and Testing. New York: Russell Sage Foundation, 1967.

TABLE 2

ANALYSIS OF COVARIANCE ON IQ CHANGE SCORES -- PRE IQ AS COVARIATE

Source	df	MS	F
Regression	1	2014.94	21.43
Treatment (A)	3	74.77	---
Sex (B)	1	332.48	3.54
Teacher Opinion (C)	2	662.18	7.04**
Socio-economic Level (D)	1	1336.80	14.22**
A X B	3	57.40	---
A X C	6	125.84	1.34
A X D	3	70.53	---
B X C	2	63.84	---
B X D	1	11.55	---
C X D	2	404.32	4.30*
A X B X C	6	64.66	---
A X B X D	3	80.99	---
A X C X D	6	157.38	1.67
B X C X D	2	137.48	1.46
A X B X C X D	6	51.57	---
Residual	810	94.03	

* $p < .05$
** $p < .01$

TABLE 3

ADJUSTED IQ CHANGE SCORE MEANS FOR MAIN EFFECTS

(Unadjusted Means in Parentheses)

EFFECT	LEVEL	N	$\bar{X}$
Treatment	Inflated IQ	233	7.32 (7.28)
	Primary Mental Abilities	206	7.39 (7.34)
	No Information	196	6.02 (6.13)
	Traditional IQ	224	6.95 (6.93)
Sex	Female	427	7.48 (7.19)
	Male	432	6.30 (6.58)
Teacher Opinion	High	305	8.08 (7.97)
	Middle	394	6.94 (6.74)
	Low	160	4.49 (5.16)
Socio-economic Level	Middle	440	8.27 (7.74)
	Low	419	5.43 (5.98)

TABLE 4

ADJUSTED IQ CHANGE SCORE MEANS FOR TEACHER OPINION BY SOCIO-ECONOMIC LEVEL INTERACTION

(Unadjusted Means in Parentheses)

GROUP		N	$\bar{X}$
Opinion	SEL		
High	Middle	129	10.98 (10.43)
High	Low	176	5.95 (6.18)
Middle	Middle	278	7.55 (7.03)
Middle	Low	116	5.46 (6.04)
Low	Middle	33	3.79 (3.30)
Low	Low	127	4.67 (5.66)

TABLE 5

ANALYSIS OF VARIANCE ON TEACHERS' IQ ACCURACY RATINGS

SOURCE	df	MS	F
Treatment (A)	2	3.12	4.03*
Sex (B)	1	3.74	4.82*
Teacher Opinion (C)	2	2.45	3.16
Socio-economic Level (D)	1	2.67	3.45
A X B	2	1.95	2.52
A X C	4	1.45	1.87
A X D	2	1.43	1.85
B X C	2	4.04	5.21**
B X D	1	.03	---
C X D	2	7.89	10.18**
A X B X C	4	.56	---
A X B X D	2	1.52	1.97
A X C X D	4	.24	---
B X C X D	2	.19	---
A X B X C X D	4	.07	---
Residual	657	.78	

* $p < .05$
** $p < .01$

ADJUSTED MEAN TO IQ CHANGE SCORES FOR INTERACTION OF TEACHER OPINION BY SOCIO-ECONOMIC LEVEL

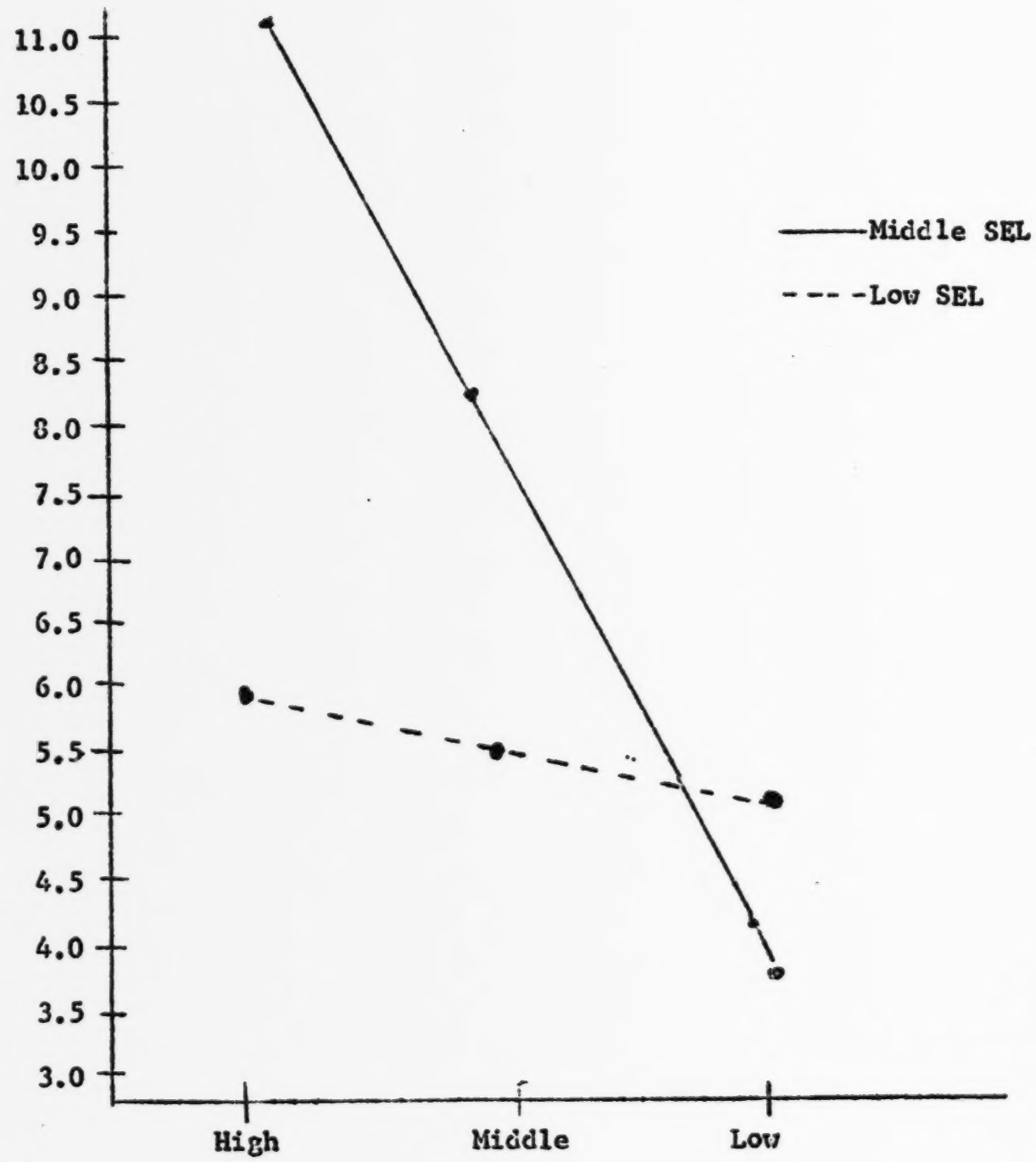

FIGURE 1